Long Range Acoustic Classification

Authors:
Ned B. Thammakhoune, Stephen W. Lang
Sanders a Lockheed Martin Company
P. O. Box 868
Nashua, New Hampshire 03061-0868

Abstract

This paper introduces the use of dynamic features for robust target recognition of ground vehicles. Most current approaches rely on instantaneous spectral features such as those derived from harmonically related spectral lines. Significant drawback of these approaches are that the use of low amplitude (10-20dB below dominant line) spectral lines severely limit classification range. The strongest line is often detectable well before secondary lines. Dynamic features extracted directly from the strongest spectral line, if successfully characterizing the target, will extend the range of operation to several times.

In this report, a complete experimental evaluation of the effectiveness of dynamic features is conducted. The analysis is performed using a database consisting of approximately two hundred acoustic signatures collected from six unique vehicles. A number of features captured from the dynamic characteristic of the spectral line are evaluated. Classification performance is measured and presented in terms of confusion matrices.

As an additional test of the classifier development tools developed for this task, we selected added instantaneous spectral measurements to the dynamic feature, and re-tested. We found that the performance of the classifiers using the mixed spectral and dynamic features was excellent, but "blind" testing of the classifiers that were developed (testing against vehicle runs that were not used during classifier development) showed disappointing results.

Introduction

The primary challenge for the success of ground vehicle classification using acoustic signature is in the area of searching for robust features for class recognition. In the past, feature design has been primarily driven by the fundamental physics of the engine mechanics, which translates acoustic energy into series of narrow band spectral peaks. These harmonically related signal components are directly related to the engine firing rate and track slap. It is then natural to classify vehicles using the feature that relate to the makeup of these harmonic lines usually detected by Harmonic Line Association (HLA) algorithm. One difficulty these techniques encounter is the low probability of detection of secondary spectral lines. It has been shown that the acoustic signature of ground vehicles is non-stationary due to many factors. Some of these dynamics are believed to be from the engine itself and some from the influence of environments such as the terrain, atmosphere and geologic characteristics. In this paper, we investigate means to extract features from the dynamic aspects of signals. The application of dynamic features in classification is motivated by the recent success of many speech recognition algorithms. Our primary objective is to evaluate classification effectiveness of transient/dynamic features that could be computed from tracking a single spectral line. If successful, it will extend the tactically useful ranges for ground vehicles several times. We used the ARL ACIDS database and a multi-variate classifier (MVG) to quantitatively evaluate our features.
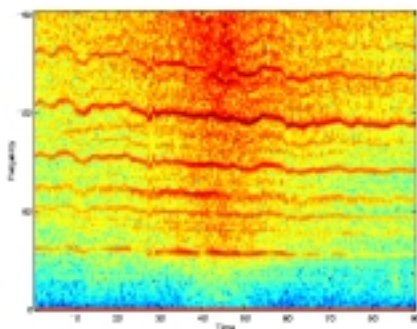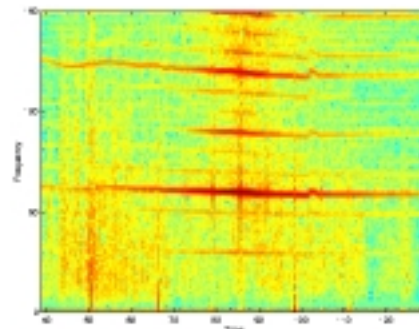

Figure 1

Figure 2

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
00 Sep 1999	N/A	-

Title and Subtitle Long Range Acoustic Classification	**Contract Number**
	Grant Number
	Program Element Number
Author(s)	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) Sanders a Lockheed Martin Company P.O. Box 868 Nashua, NH 03061-0868	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es) Department of the Army, CECOM RDEC Night Vision & Electronic Sensors Directorate AMSEL-RD-NV-D 10221 Burbeck Road Ft. Belvoir, VA 22060-5806	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)

Distribution/Availability Statement
Approved for public release, distribution unlimited

Supplementary Notes
See also ADM201471, Papers from the Meeting of the MSS Specialty Group on Battlefield Acoustic and Seismic Sensing, Magnetic and Electric Field Sensors (2001) Held in Applied Physics Lab, Johns Hopkins Univ, Laurel, MD. on 24-26 Oct. 2001. Volume 2 (Also includes 1999 and 2000 Meetings), The original document contains color images.

Abstract

Subject Terms

Report Classification unclassified	Classification of this page unclassified
Classification of Abstract unclassified	**Limitation of Abstract** UU

Number of Pages
6

Feature Design

The primary signal space we extracted feature from is the time-frequency distribution. We examined both the short time Fourier transform (STFT) and the reduced interference (RID) time-frequency distributions. The RID distribution produces better spectrum resolution as compared to the STFT distribution. It utilizes a single-side spectrum of real input signal by applying Hilbert transform. This effectively doubles the frequency resolution.   In addition, it reduces the cross interference among closely space spectral peaks by the smoothing effect of exponential kernels. It, however, introduces significant amplitude distortion. In our application, for features that depend only on the variation of the maximum frequency bin, we used the RID distribution to capture more details of spectral variation. For features that depend on the amplitude, we used the STFT distribution as the feature's signal space. We focused mostly on means of measuring the time evolving characteristics of the strongest spectral line. Figures1 and Figure 2 shows examples of the RID distribution of two different vehicles under the same driving environments. Clearly, it illustrates different rate of change for the maximum frequency of the strongest spectral line. The images in figure 1 and 2were locally normalized to enhance the spectral line over the time scale. It is also important to note that all the spectral lines share the same dynamic characteristics over time; thus it is adequate to capture dynamic behavior from one single line without any loss of information. A list of the features that we extracted is shown in Table 1.

Standard deviation of F_{max}
Number of positive dF_{max}/dt
Standard deviation of dF_{max}/dt
Standard deviation of dA_{max}/dt over dF_{max}/dt
Sum of dF_{max}/dt
Number of zero crossing of dF_{max}/dt
Sum of $dF_{max}/dt/F_{max}$ over F_{max}

Table 1

Feature Extraction and Optimization

This section briefly describes how we systematically extracted features from the acoustic signature. We first removed DC bias by performing trend removal. We then calculated STFT and RID time-frequency distributions. The frame size is set to 1 seconds using 50 percent overlap. Based on the signal to noise ratio, we tracked the strongest spectral line and extracted maximum frequency bin versus time (F(t)). From the tracked spectral line, we then captured all the features of interest. Following that, we associated each feature vector with type of vehicle, environment and speed using ground truth. We performed a quick analysis of each feature by inspecting the probability density distributions (pdf). Figure 3 and 4 show examples of pdf's of two features. As depicted, class separation is obvious among some classes while others exhibit considerable overlap. The pdf 's also approximate Gaussian distribution to some degree. The pdf analysis gave us an early indication that this is a complex class boundary problem. We then considered feature analysis that accommodates for the feature correlation. We chose a sub-optimal multidimensional feature ranking technique to perform further feature analysis. Ideally, we would prefer the exhaustive search method in which every M out of N feature combinations are tried for the best performance. However, because the number of combination increases prohibitively with the number of features, the implementation is impractical. We thus resort to a sub-optimal search method known as "add-on" to find a
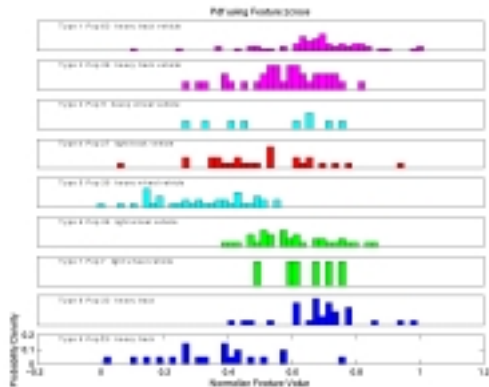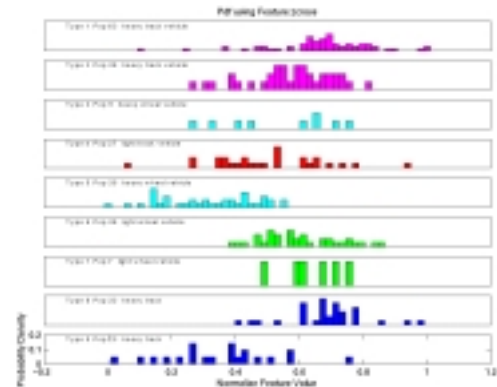

Figure 3

Figure 4

reasonably good feature subset. The algorithm first evaluates classification performance of each of the N features independently and selects a single best feature. It then proceeds to evaluate performance of the next N-1 two-feature subset, and selects the best. The process repeats in the same manner, each time adding the one feature that maximizes the performance. This method then evaluates M(2N+1-M)/2 subset to reach the best M-feature subset.

Vehicle Type	Classifier Output									%
1 Heavy Track Vehicle	47	6	0	0	1	3	0	4	0	0.77
2 Heavy Track Vehicle	9	16	0	0	4	4	0	1	2	0.44
3 Heavy Wheel Vehicle	6	0	0	0	0	2	0	1	0	0.00
4 Heavy Track Vehicle	6	5	0	3	7	0	0	0	6	0.11
5 Heavy Wheel Vehicle	6	2	0	0	21	1	0	1	8	0.54
6 Heavy Wheel Vehicle	3	0	0	0	0	27	0	2	4	0.75
7 Heavy Wheel Vehicle	2	0	0	0	0	1	0	3	0	0.00
8 Heavy Track Vehicle	16	0	0	0	0	1	0	15	1	0.45
9 Heavy Track Vehicle	0	2	0	0	6	6	0	0	7	0.33

Table 2

Classification Performance Analysis

To evaluate the target recognition performance of the optimized feature set, we generated a classification performance ROC. Because of the limited number of target signatures we have for each class, we had to train and test the classifier using the single hold out method to maximize the training set. This minimizes the error due to under-training. We chose the classical Multi-variate Gaussian Classifier as the primary classifier for this analysis. We also performed the same analysis using PNN and NNC classifiers for comparison purposes. Multivariate Gaussian Classifier (MVG) is a classical conventional classifier that assumes a Gaussian distribution of underlying features. It parameterizes each class mean and covariance matrix and classifies by minimizing the nearest mean. Its performance degrades if the assumed models are mismatched. The Probabilistic Neural Network (PNN), on the other hand, is a non-parametric neural network classifier that makes no assumptions on the underlying feature distribution. It utilizes a Gaussian kernel function with a smoothing coefficient as activation function for neurons and classifies by summing feature vector distance from all training data. Its performance degrades if the training data are limited.

Table 2 shows the result of target identification for all 9 vehicles. The recognition percentage for vehicle1 and vehicle 6 are among the highest score at 70's. Vehicle 2,5 8,9 scored ranging from 33 to 54 %. For vehicle 3,4 and 7, the very low scores reflected the fact that there were very small number of acoustic signatures for the class to be properly trained. We grouped the vehicles of same definition together and performed the same classification analysis. The result is shown in Table 3. Similar results were obtained. Again, class 2 scores the lowest because of the small population in its class.

	Classifier Output					%
1 Heavy track vehicle	61	8	4	24	0	**0.63**
2 Heavy wheel vehicle	9	9	0	9	0	**0.00**
3 Light track vehicle	4	23	2	18	1	**0.48**
4 Light wheel vehicle	5	2	23	12	0	**0.54**
5 Heavy track	10	5	5	34	0	**0.63**

Table 3

	Output	%
Heavy	174	25	**0.87**
Light	41	28	**0.41**

Table 5

	Output	%
Track	129	49	**0.72**
Wheel	26	71	**0.71**

Table 4

Combined Spectral and Dynamic Features
In this part of the effort, we combined traditional spectral features with the dynamic features described above. The complete list of features is provided in Table 6.

Frequency of loudest tone
Ratio of (frequency of second loudest tone)/frequency of loudest tone
Ratio of (frequency of third loudest tone)/frequency of loudest tone
Ratio of (frequency of third loudest tone)/frequency of second loudest
Ratio of (power of second loudest tone)/power of loudest tone
Ratio of (power of third loudest tone)/power of loudest tone
Ratio of (power of third loudest tone)/power of second loudest
Number of zero crossing of dF_{max}/dt (20 second window) (loudest tone)
Sum of dF_{max}/dt (20 second window) (loudest tone)
Standard deviation of dF_{max}/dt (20 second window) (loudest tone)
Number of zero crossing of dF_{max}/dt (7 second window) (loudest tone)
Sum of dF_{max}/dt (7 second window) (loudest tone)
Number of zero crossing of dF_{max}/dt (7 second window) (loudest tone)
Number of zero crossing of dF_{max}/dt (20 second window) (second loudest tone)
Sum of dF_{max}/dt (20 second window) (second loudest tone)
Standard deviation of dF_{max}/dt (20 second window) (second loudest tone)
Number of zero crossing of dF_{max}/dt (7 second window) (second loudest tone)
Sum of dF_{max}/dt (7 second window) (second loudest tone)
Standard deviation of dF_{max}/dt (7 second window) (second loudest tone)
Ratio of frequency of loudest seismic tone/loudest acoustic tone
Ratio of power in lowest seismic tone/power in loudest seismic tone
Ratio of frequency of lowest seismic tone/frequency of loudest seismic tone
Number of seismic tones that match acoustic tones in frequency
ratio of frequency of lowest acoustic tone/loudest acoustic tone
Ratio of power in lowest acoustic tone/power in loudest acoustic tone
ratio of frequency of lowest (harmonic) tone/loudest acoustic tone
Number of acoustic tones in target
Number of seismic tones in target
frequency of loud harmonic/frequency of loud tone
power of loud harmonic/power of loud tone
power of low frequency harmonic/power of loud tone
frequency of loudest seismic tone
frequency of loud harmonic
frequency of low harmonic
instantaneous spectral width of loudest tone
average spectral width of loudest tone
variance of the spectral width of loudest tone
instantaneous spectral width of second loudest tone
average spectral width of second loudest tone

variance of the spectral width of loudest tone
ratio of spectral width of the loudest and second loudest tones
ratio of average spectral width of the loudest and second loudest tones
mean of the absolute value of dF/dt for loudest tone
Total acoustic power in the 0 – 100 Hz band in the direction of the target
Total acoustic power in the 100-200 Hz band in the direction of the target
Broadband acoustic power in the 0 – 100 Hz band in the direction of the target (tones excluded)
Broadband acoustic power in the 100-200 Hz band in the direction of the target (tones excluded)
Total acoustic power in the 0 – 67 Hz band in the direction of the target
Total acoustic power in the 67-132 Hz band in the direction of the target

Total acoustic power in the 132-200 Hz band in the direction of the target
Broadband acoustic power in the 0 – 67 Hz band in the direction of the target (tones excluded)
Broadband acoustic power in the 67-132 Hz band in the direction of the target (tones excluded)
Broadband acoustic power in the 132-200 Hz band in the direction of the target (tones excluded)
Fundamental Frequency of the loudest harmonic set
Acoustic Power level of the first 8 harmonics of the set (normalized by power of the loudest tone)
Ordered Harmonic numbers of the loudest 3 harmonics
Fundamental Frequency of the loudest harmonic set (Alternate fundamental estimation technique)
Acoustic Power level of the first 8 harmonics of the set (alternate technique)
Ordered Harmonic numbers of the loudest 3 harmonics (alternate technique)
Number of harmonic sets detected

Table 6

Since we wished to test the utility of seismic features, and we did not have the seismic portion of the ACIDS database, we switched to using our own database, with a small number of target runs collected at Aberdeen in December 1998, and at Fort Irwin in February 1999.

The tools described earlier were used to analyze these features, and to rank them in terms of their utility as classification features. The initial run showed that the frequency information (Frequency of the loudest tone and fundamental frequency of the loudest harmonic set were the most valuable features available. After considering this result, we decided that we had only a small number of target runs, with a limited number of vehicle speeds, so our sampling of frequencies was too limited, to use as a classifier input.

After excluding the two frequency features, we re-ran the analysis and found that the seismic-related features (Number of seismic tones that match acoustic tones in frequency, Ratio of frequency of loudest seismic tone/loudest acoustic tone, Ratio of power in lowest seismic tone/power in loudest seismic tone
Ratio of frequency of lowest seismic tone/frequency of loudest seismic tone, Number of seismic tones that match acoustic tones in frequency, Number of seismic tones in target) were among the top-ranked features. After closer examination, we found that the hardware configuration for the seismic sensor changed dramatically between the Aberdeen and Irwin data collection exercises, and the classifiers were using this difference to distinguish between the US vehicles collected at Aberdeen and the Soviet vehicles from Ft. Irwin. After failing to find a method to compensate for the hardware changes, we decided to exclude these features from subsequent analyses.

The final analysis, with the feature set now pruned to include only the reliable features, yielded a short list of features that are most valuable for classification

Ratio of the frequency of the second loudest tone to the loudest tone
Ratio of the powers of the second loudest and loudest tones
Mean dF/dt for the second loudest tone (7 second window)
Average width of the second loudest tone
Mean dF/dt for the loudest tone
Number of acoustic tones detected
Average spectral width of the loudest tone
Variance of the spectral width of the loudest tone

With these 8 features, the vehicle ID performance was about 75% correct. A blind test was performed using a few runs that were excluded from the data sets used to develop the classifier. The blind test showed that the classifier performance was only about 55% correct. From this, we conclude that the number of vehicle runs in the target database was insufficient to develop a reliable classifier (average of 3 pass-by's per vehicle type).

A final test was performed using just the relative power of the first 8 harmonics of the loudest harmonic set that was detected. Using these 8 features, the classifier performance against the train/test set was only about 55%. The performance on the blind set, however, was also 55% correct, from which we conclude that these features are robust in the face of a small training set.

Summary

The search for robust features will continue to be an important area of target recognition for ground vehicles. Different aspects of signals should be exploited to extract many uncorrelated features for versatility, and effectiveness. In this report, our preliminary investigation[1] shows moderate success of using dynamic features alone in target ID for different class category partition. It is less likely that these features are highly correlated with HLA based features simply because of the way they were extracted. This suggests the possibility of performance improvement when the two feature sets are combined and optimized for the best combination subset. In the future, a more complicated method of extracting dynamic feature should be studied.